

VAI TRÒ CỦA KHOA HỌC DỮ LIỆU TRONG NỀN KINH TẾ SỐ

● NGUYỄN THỊ PHƯƠNG DUNG

TÓM TẮT:

Bài viết mô tả vai trò của khoa học dữ liệu và dữ liệu lớn (Big data) trong thời đại số hóa. Các công ty trong các lĩnh vực thương mại và dịch vụ bán lẻ rất quan tâm tới việc sử dụng dữ liệu lớn và công nghệ số. Các công nghệ này được các ngân hàng, các công ty sử dụng để phân tích dữ liệu lớn, nhằm tìm ra các lỗi thiết bị, giảm thời gian chết, từ đó giảm chi phí hoạt động. Công nghệ dữ liệu lớn được xem như là chất xúc tác của quá trình sản xuất và là điều kiện cần thiết để gia tăng lợi nhuận của doanh nghiệp thông qua dịch vụ khách hàng được cá nhân hóa và phân tích dự báo.

Từ khóa: công nghệ dữ liệu lớn, kinh tế số.

1. Đặt vấn đề

Chúng ta đang sống trong thời đại “chuyển đổi số”. Các thiết bị điện tử như máy tính, điện thoại và nhiều loại thiết bị khác đã trở nên rất rẻ và rất phổ biến. Ngày nay công nghệ đang thay đổi rất nhanh chóng và động lực chính của những thay đổi này chính là “số hóa khách hàng”. Tốc độ thay đổi sở thích của khách hàng làm gia tăng nhu cầu về chất lượng, tính năng và thiết kế của các loại sản phẩm. Mô hình tiêu dùng đang thay đổi, từ đó nổi lên xu hướng rất rõ ràng: tương tác thông qua các thiết bị thông minh. Nếu công nghệ mà lỗi thời thì không thể mang lại lợi nhuận bền vững trong dài hạn, vì việc hiện đại hóa của chúng sẽ rất tốn kém và đòi hỏi tài nguyên tương đối lớn để bảo trì. Thị trường toàn cầu cho dữ liệu lớn đang được phát triển rất nhanh, đặc biệt trong lĩnh vực công, y tế, sản xuất và bán lẻ.

Trong năm 2018, người dùng và doanh nghiệp trên toàn thế giới đã tạo ra 33 zettabyte thông tin,

năm 2025 sẽ có 175 zettabyte. Động lực tăng trưởng chính là các nền tảng giải trí, hệ thống giám sát video, thiết bị kết nối internet, các dụng cụ làm gia tăng năng suất. Tất cả những điều này đều được làm việc với dữ liệu lớn. Chuyển đổi số làm gia tăng tốc độ và tính linh hoạt của việc tạo ra giá trị, cho phép thích ứng với tốc độ cao của việc thay đổi thị trường và các yêu cầu của người tiêu dùng thời đại kỹ thuật số. Chính điều này đòi hỏi làm tăng yêu cầu về năng lực đối với các nhân viên của mỗi công ty, tăng khả năng làm việc với công nghệ của họ, cùng với cách tiếp cận và phương pháp mới trong quản lý dự án, phát triển và vận hành. Điều này cũng liên quan tới các hình thức quản lý linh hoạt, nâng cao chất lượng tương tác của truyền thông.

2. Kinh tế số và khoa học dữ liệu là gì?

Kinh tế số: Nhóm cộng tác kinh tế số Oxford đã đưa ra khái niệm về kinh tế số, theo đó kinh tế số được hiểu là một nền kinh tế vận hành chủ yếu dựa

trên nền công nghệ số, trong đó các giao dịch mua bán là giao dịch điện tử được tiến hành thông qua Internet. Kinh tế số bao gồm tất cả các lĩnh vực của nền kinh tế (công nghiệp, nông nghiệp, dịch vụ; sản xuất, phân phối, lưu thông hàng hóa, giao thông vận tải, logistic, tài chính ngân hàng, ...).

Về bản chất, kinh tế số là các mô hình tổ chức và phương thức hoạt động của nền kinh tế dựa trên các ứng dụng công nghệ số. Ví dụ như: các trang thương mại điện tử, quảng cáo trực tuyến hay các ứng dụng về ăn uống, vận chuyển, giao nhận,... Xét ở tầm vĩ mô, kinh tế số có những đóng góp không nhỏ trong sự hội nhập của các doanh nghiệp vào chuỗi công nghệ toàn cầu và tạo ra các giá trị về kinh tế lớn, thúc đẩy phát triển đất nước.

Khoa học dữ liệu: là khoa học về việc quản trị và phân tích dữ liệu để tìm ra các hiểu biết, các tri thức hành động, các quyết định dẫn dắt hành động. Khoa học dữ liệu gồm ba phần chính: Tạo ra và quản trị dữ liệu, phân tích dữ liệu, và chuyển kết quả phân tích thành giá trị của hành động. Việc phân tích và dùng dữ liệu lại dựa vào ba nguồn tri thức: toán học (thống kê toán học), công nghệ thông tin (máy học) và tri thức của lĩnh vực ứng dụng cụ thể. Khoa học dữ liệu còn là lĩnh vực nghiên cứu dữ liệu nhằm khai thác những thông tin chuyên sâu có ý nghĩa đối với hoạt động kinh doanh.

Khoa học dữ liệu rất quan trọng đối với nền kinh tế bởi vì lĩnh vực này kết hợp các công cụ, phương pháp và công nghệ để rút ra ý nghĩa từ dữ liệu. Khoa học dữ liệu đang cách mạng hóa phương thức hoạt động của các doanh nghiệp, bất kể quy mô bởi các doanh nghiệp này đều cần một chiến lược khoa học dữ liệu nhằm thúc đẩy tăng trưởng và duy trì lợi thế cạnh tranh của mình.

3. Vai trò của khoa học dữ liệu trong nền kinh tế số

3.1. Cơ sở lý thuyết

Công nghệ mới và dữ liệu lớn đang tác động rất lớn đến việc thay đổi các hoạt động trong doanh nghiệp, nó không chỉ tác động vào các vị trí

các lãnh đạo, mà còn thay đổi liên tục việc đào tạo nhân viên ở tất cả các cấp độ. Khoa học dữ liệu đang làm thay đổi và các doanh nghiệp đều đang dần thích nghi với việc triển khai các dự án về lĩnh vực khoa học dữ liệu, dù đây là lĩnh vực không phải lúc nào cũng mang lại lợi ích thương mại. Người đứng đầu chỉ đạo triển khai khoa học dữ liệu trong doanh nghiệp (Chief Data Office - CDO) phải đối mặt với nhiều vai trò, như: chuyên gia về công nghệ tiên tiến, góp ý vào chiến lược kinh doanh của doanh nghiệp, và đưa ra các ý tưởng liên quan đến sử dụng dữ liệu. Đồng thời, các CDO cũng phải chịu trách nhiệm về chất lượng và tính minh bạch của dữ liệu do doanh nghiệp cung cấp trước các cơ quan quản lý. Trách nhiệm của các CDO là xây dựng một chiến lược về phát triển năng lực phân tích dữ liệu của doanh nghiệp và các kỹ năng cần phát triển của nhân viên, kiểm soát hệ thống công nghệ của tổ chức, trực tiếp quản lý và cho phép trích xuất và xử lý trực quan hóa dữ liệu nhằm đem đến hiệu quả tối ưu nhất.

Hiện nay hầu hết các tổ chức, doanh nghiệp đều thiếu nhân lực có kỹ thuật cao trong lĩnh vực khoa học dữ liệu bởi các chuyên gia phân tích dữ liệu giỏi thường có xu hướng tham gia làm việc ở các ngành công nghệ thông tin. Trước đây, các máy tính lớn hầu như chỉ có ở các doanh nghiệp, nhưng hiện nay, đa phần người dùng đều cần một máy tính xách tay để tìm kiếm thông tin và xử lý những vấn đề mà mình quan tâm với hiệu quả cao nhất. Thế giới đang hướng tới nguồn dữ liệu mở, các gói dữ liệu dùng cho phân tích có thể lấy xuống miễn phí và hợp pháp. Bên cạnh đó, các chương trình giáo dục từ các tổ chức thương mại và phi lợi nhuận luôn được mở miễn phí cho người dùng nếu có thời gian và nhu cầu học tập.

Văn hóa của một doanh nghiệp sẽ được hình thành trước khi có sự phổ biến rộng rãi của khoa học dữ liệu, đó là cách tiếp cận rất thận trọng với các rủi ro, giảm thiểu chi phí và cải thiện sự ổn định của hoạt động doanh nghiệp. Trước đây, các

công ty chỉ cố gắng sử dụng các mô hình kinh doanh và công nghệ đã được thử nghiệm qua thời gian, năng lực tài chính thường được thuê ngoài để tiết kiệm tiền bạc. Điều này xảy ra bởi lẽ các thị trường đã được hình thành trước đó. Doanh nghiệp không thể thực hiện một cuộc chuyển đổi quy mô lớn bằng cách gửi từng cá nhân đơn lẻ tham gia các khóa đào tạo lại. Để làm được điều này, đòi hỏi nhiều cách tiếp cận cơ bản và triệt để hơn, yêu cầu sự tham gia và hỗ trợ sâu hơn từ những người đứng đầu doanh nghiệp, người giám sát tất cả các khía cạnh của quản lý nhân sự và chính sách quan hệ, thực hành và hoạt động cho tổ chức công nghệ cao. Sử dụng tất cả các kỹ năng và khả năng mà nguồn nhân lực có, họ sẽ giúp cho doanh nghiệp tạo ra sự cân bằng từ đầu tới cuối chương trình phát triển, giải quyết được tất cả các thách thức.

Một trong những ví dụ để giải quyết vấn đề này là tạo ra nhân viên lao động “dù” phát triển chương trình theo hướng dữ liệu lớn bằng cách:

- Hàng năm, doanh nghiệp sẽ đào tạo những nhà quản lý hàng đầu, có thể tìm hiểu về khả năng của khoa học dữ liệu và các trường hợp thành công từ khắp nơi trên thế giới.

- Đào tạo cơ bản cho các nhà quản lý trung cấp nhằm cho họ thấy những lợi ích của việc sử dụng khoa học dữ liệu và công cụ học máy cho các loại nhiệm vụ khác nhau.

- Cử các nhà quản lý và khoa học dữ liệu của doanh nghiệp tham gia các cuộc thi lập trình thường xuyên, nhằm tương tác, trao đổi với người đồng cấp ở các doanh nghiệp khác.

- Thường xuyên tổ chức các cuộc thi trực tuyến cho những người có sở hữu các tập dữ liệu lớn.

- Cử các nhà khoa học dữ liệu của doanh nghiệp tham gia đào tạo kỹ năng máy tính tiên tiến hàng năm.

- Lên các chương trình đào tạo cá nhân cho các nhà quản lý và nhà khoa học dữ liệu của doanh nghiệp.

3.2. Phương pháp luận

Mục tiêu phát triển dự án khoa học dữ liệu hay phát triển phần mềm đều nhằm thu được lợi ích kinh tế hoặc tạo ra thêm giá trị cho khách hàng.

Một dự án trong khoa học dữ liệu tương tự như phát triển phần mềm. Trong cả hai trường hợp thì mục tiêu đều là đạt được lợi ích kinh tế hoặc gia tăng giá trị cho khách hàng. Để đạt được nó, chúng ta sẽ phải viết “code”. Đó là lý do tại sao các nhà quản lý thường mong đợi những kết quả từ việc xử lý dữ liệu khoa học như trong việc phát triển phần mềm và thành công được đo lường chính xác bởi những mục tiêu mà họ mong muốn đã đạt được. Một trong những cách tiếp cận phổ biến nhất đối với phân tích dữ liệu là xử lý liên ngành để khai thác dữ liệu, ở đó mỗi tác vụ được trình bày dưới dạng giả thuyết nào đó, có thể được chấp nhận hay bác bỏ. Tất cả các việc làm trên các giả thuyết đó qua rất nhiều giai đoạn: đánh giá tập dữ liệu cần thiết, xác định các tiêu chí thành công và chất lượng của số liệu, mô hình hoá và đánh giá kết quả. Thông thường, công việc sẽ diễn ra theo một chu kỳ, một số giai đoạn có thể được lặp đi lặp lại nhiều lần. Nếu ta thấy mô hình đã tạo ra được những cải thiện quy trình hiện tại và sản phẩm của công ty thì mô hình đó sẽ được đưa vào hoạt động. Tuy nhiên, cũng có thể mô hình đó không mang lại hiệu quả của việc cải thiện. Nếu tình huống này gặp phải, ta không coi đó là một kết quả tiêu cực, vì nhóm dự án đã thu được rất nhiều kinh nghiệm và kiến thức, có thể được tiếp tục áp dụng vào các giả thuyết khác. Cách tiếp cận phân tích dữ liệu liên ngành là một quy trình tiêu chuẩn để kiểm tra dữ liệu. Mô hình để phân tích dữ liệu bao gồm 6 giai đoạn, trình tự của các giai đoạn không xác định nghiêm ngặt. Theo quy định, trong hầu hết các dự án, người thực hiện thường phải quay trở về và kiểm tra các giai đoạn trước.

Các giai đoạn của một vòng nghiên cứu dữ liệu được mô tả bởi các giai đoạn sau đây:

1. Hiểu biết về kinh doanh bao gồm: xác định mục tiêu kinh doanh; đánh giá tình huống kinh

doanh; xác định mục tiêu của phân tích dữ liệu; lập kế hoạch cho dự án.

2. Hiểu dữ liệu: thu thập dữ liệu; mô tả dữ liệu; nghiên cứu dữ liệu; kiểm tra tính đúng đắn của dữ liệu.

3. Chuẩn bị dữ liệu nhằm mục đích để: lựa chọn kiểu dữ liệu; làm sạch dữ liệu bao gồm chuyển đổi dữ liệu và chuẩn bị mẫu; dẫn xuất dữ liệu; hợp nhất dữ liệu và tùy chỉnh dữ liệu.

4. Mô hình cần dùng: lựa chọn phương pháp dùng cho mô hình; thực hiện 1 thử nghiệm cho mô hình; xây dựng mô hình; đánh giá mô hình.

5. Đánh giá: đánh giá kết quả; xem xét quá trình; xác định các bước tiếp theo.

6. Triển khai: Lên kế hoạch triển khai; lập kế hoạch hỗ trợ và giám sát việc triển khai; làm báo cáo cuối cùng; đánh giá quá trình.

Ví dụ: Trong trường hợp một ngân hàng, một mô hình thành công có thể được minh họa bằng cách dự đoán các chỉ số gián tiếp và dữ liệu lịch sử về thu nhập của khách hàng. Trên cơ sở dữ liệu thu được, ngân hàng có thể cung cấp một sản phẩm cho vay cá nhân mà không cần xác định thu nhập chính thức của khách hàng.

3.3. Thảo luận

Ta thấy, dữ liệu là cụm từ quan trọng nhất gắn liền với thuật ngữ “kinh tế số”. Chu kỳ mới của một nền kinh tế dựa trên việc thu thập dữ liệu, chất lượng của dữ liệu và phân tích dữ liệu. Điều này là rất khả thi và phù hợp với nhu cầu của các thành phần kinh tế khác nhau, để hỗ trợ các giải pháp của họ ở khắp mọi nơi. Tất cả các doanh nghiệp, người tiêu dùng luôn quan tâm tới dữ liệu kỹ thuật số vì nó rất hữu ích, thuận tiện và chuyên nghiệp trong xử lý các thông tin cần đạt tới. Hiệu suất của công việc được cải thiện rất nhiều với các công nghệ thông tin kỹ thuật số hiện nay.

Dữ liệu cho phép số hóa thực trạng kinh doanh xung quanh chúng ta và xây dựng các mô hình cấp cao để hiểu sâu hơn việc phân tích kinh doanh hoặc dự đoán phân tích nhanh khi mà thời gian đưa ra

quyết định là rất ngắn. Dữ liệu mở ra các khía cạnh mới của các sự kiện và hiện tượng giúp cho doanh nghiệp đưa ra quyết định chấp nhận hay bác bỏ các giả thuyết của các nhà lý thuyết. Đối lại, dữ liệu kỹ thuật số buộc nền kinh tế phải chịu chi phí đáng kể để xây dựng cơ sở hạ tầng phục vụ cho việc sử dụng dữ liệu. Dữ liệu trung tâm, mạng truyền thông tốc độ cao... là các khía cạnh kỹ thuật của toàn cầu hóa dữ liệu.

Dữ liệu buộc các doanh nghiệp phải suy nghĩ thận trọng hơn về phương thức và cách sử dụng dữ liệu để đem lại hiệu quả sử dụng cao nhất cho mục đích chính của doanh nghiệp. Kết quả là một số lượng lớn các dự án sáng tạo, các nhóm làm việc có phương pháp dần dần tạo ra dữ liệu và xử lý dữ liệu, trích xuất ra những giá trị bổ sung từ dữ liệu ban đầu. Về cơ bản cần có công cụ và các cách tiếp cận dữ liệu khác nhau. Tuy nhiên, để trích xuất dữ liệu đúng cách và làm việc với dữ liệu một cách định tính, thì ta cần phải thực sự hiểu về nó. Chẳng hạn, với các nhà tiếp thị, việc thực sự đi vào trái tim của khách hàng phải thực sự đi sâu vào các chi tiết khác nhau và sự phức tạp của các giao tiếp xã hội.

4. Kết quả

Trước sự bùng nổ của cách mạng công nghệ 4.0, khoa học dữ liệu đang ngày càng trở nên quan trọng đối với các doanh nghiệp. Do vậy, các doanh nghiệp ở các quy mô và các ngành nghề ngày càng chủ động quan tâm và hướng tới các công nghệ dữ liệu lớn, nhất là các doanh nghiệp trong lĩnh vực tài chính, ngân hàng và các nhà khai thác di động. Ngoài ra, khoa học dữ liệu còn được các doanh nghiệp sản xuất lớn sử dụng để phân tích dữ liệu về sự cố thiết bị, nhằm giảm thời gian ngừng hoạt động, từ đó giúp giảm chi phí hoạt động doanh nghiệp. Ví dụ, trong lĩnh vực quản lý chuyến bay, việc phân tích các bộ dữ liệu có thể tăng độ tin cậy của thiết bị và giảm số lần hỏng hóc.

Nhưng lĩnh vực ứng dụng của khoa học dữ liệu rộng hơn nhiều. Ví dụ, một lĩnh vực ứng dụng khác của công nghệ dữ liệu lớn có thể là mua một doanh nghiệp, do vậy, nếu bao quát được càng nhiều khía

cạnh khác nhau của một công ty thì hoạt động phân tích dữ liệu càng phù hợp với thực tế. Trình tự phân tích dữ liệu theo thứ tự sau: 1) Các tính năng tài sản vật chất; 3) Dữ liệu thao tác; 3) Dữ liệu tài chính; 4) Tài nguyên về nguyên vật liệu; 5) Dữ liệu pháp lý.

Các nguồn thông tin để xử lý có thể là: bảng tệp, hệ thống quản lý cơ sở dữ liệu truyền thống và hệ thống kế toán. Cũng cần phân tích các dữ liệu hỗn độn như: phản hồi của khách hàng, kết quả giám định, những yêu cầu về dịch vụ hay sự cố, môi trường cạnh tranh và hạ tầng công nghệ thông tin. Ở đây các nguồn thông tin có thể là: bảng, sơ đồ, mạng xã hội, đánh giá của chuyên gia,... Cái gọi là hộ chiếu đối tượng, bao gồm thông tin về vị trí, diện tích, số tầng, giấy phép, dữ liệu hàng tồn kho, môi trường cạnh tranh, dữ liệu tài chính lịch sử, yếu tố bán hàng theo mùa,..., có thể là kết quả cuối cùng của việc phân tích dữ liệu đó.

Tuy nhiên, như đã được công nhận, giá trị của công ty không giống như tổng số tài sản hữu hình đơn giản của doanh nghiệp, việc thu thập và phân tích dữ liệu trên công nghệ dữ liệu lớn sẽ giúp doanh nghiệp ước tính được giá trị của tài sản vô hình của mình. Các tài sản vô hình của doanh nghiệp có thể bao gồm: 1) Nguồn nhân lực, kiến thức và kỹ năng; 2) Nguồn thông tin, cơ sở dữ liệu; cơ cấu tổ chức quản lý; 3) Nhân viên tiềm năng và tài giỏi; 4) Thương hiệu và danh tiếng; 5) Sự phát triển; 6) Cơ sở khách hàng; 7) Mối quan hệ với các đối tác.

Hiện nay, các doanh nghiệp nước ngoài có quy mô vừa và lớn đều có các CDO tiêu chuẩn, công việc của các CDO là làm việc trực tiếp với các quản lý cấp cao. Tuy nhiên, ở Việt Nam, vị trí làm việc này còn rất ít, các chuyên gia về phân tích và quản

lý dữ liệu lớn còn rất thiếu và cần được đào tạo lại. Bên cạnh đó, các chuyên gia giỏi yêu cầu rất cao về mức lương khởi điểm và còn bị hạn chế khả năng làm việc bởi sự thiếu hiểu biết của đa số lãnh đạo doanh nghiệp trong lĩnh vực quản lý dữ liệu lớn. Do vậy, Nhà nước cần có chính sách trong việc chuyển đổi nền kinh tế số để các lãnh đạo doanh nghiệp hiểu được tầm quan trọng của việc quản lý hiệu quả dữ liệu lớn.

5. Kết luận

Ở thời điểm hiện tại, những doanh nghiệp tư nhân hay nhà nước đều quan tâm đặc biệt tới việc quản lý dữ liệu có cấu trúc và phi cấu trúc, đồng hành cùng là sự hỗ trợ của các loại công cụ và công nghệ mới. Nhờ phân tích dữ liệu lớn (tức là các phương pháp xử lý lượng lớn dữ liệu cho các nhiệm vụ và mục đích cụ thể), khả năng phân tích được mở rộng đáng kể và có thể thu thập được thông tin giá trị cho khách hàng. Tận dụng được năng lực tiềm tàng của dữ liệu lớn, các công ty có thể nhận được nhiều thông tin hữu ích về đối thủ cạnh tranh, đối tác và khách hàng. Các công nghệ dữ liệu lớn không còn là “kho báu quốc gia” trong thời đại của các tập đoàn đa quốc gia và sự hợp tác trên toàn thế giới nữa. Vai trò của dữ liệu lớn giờ đây như một loại chất lỏng, là điều kiện tiên quyết cho các tổ chức cải thiện được lợi nhuận, thông qua dịch vụ khách hàng được cá nhân hóa và phân tích dự đoán.

Đối với nền kinh tế kỹ thuật số của các nước nói chung, điều cần thiết là đạt được việc thiết lập trao đổi dữ liệu cũng như hợp pháp hóa một định nghĩa duy nhất về dữ liệu lớn. Đây sẽ là một thành phần quan trọng trong khả năng cạnh tranh của nền kinh tế trên thị trường toàn cầu và là một bước tiến lớn trong việc thúc đẩy doanh nghiệp trong nước ■

TÀI LIỆU THAM KHẢO:

1. F. Buytendijk (2014). Hype Cycle for Big Data. USA: Gartner Research.
2. Tukey, J. (1977). Exploratory Data Analysis. New York: Addison-Wesley.

3. Franks, B. (2012). Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. USA: John Wiley & Sons.

Ngày nhận bài: 15/2/2023
Ngày phản biện đánh giá và sửa chữa: 15/3/2023
Ngày chấp nhận đăng bài: 26/3/2023

Thông tin tác giả:
NGUYỄN THỊ PHƯƠNG DUNG
Học viện Ngân hàng

THE ROLE OF DATA SCIENCE
IN THE DIGITAL AGE

● NGUYEN THI PHUONG DUNG
Banking Academy

ABSTRACT:

This paper highlights the role of data science and Big data in the digital age. Many commerce and retail companies are paying a great attention to data analysis and digital technologies. Data and digital technologies have been used by banks and telecommunication companies to analyze data in order to reduce downtime and costs. Big data is considered a catalyst for improving profitability of businesses through personalized customer service and predictive analytics.

Keywords: Big data, digital economy.